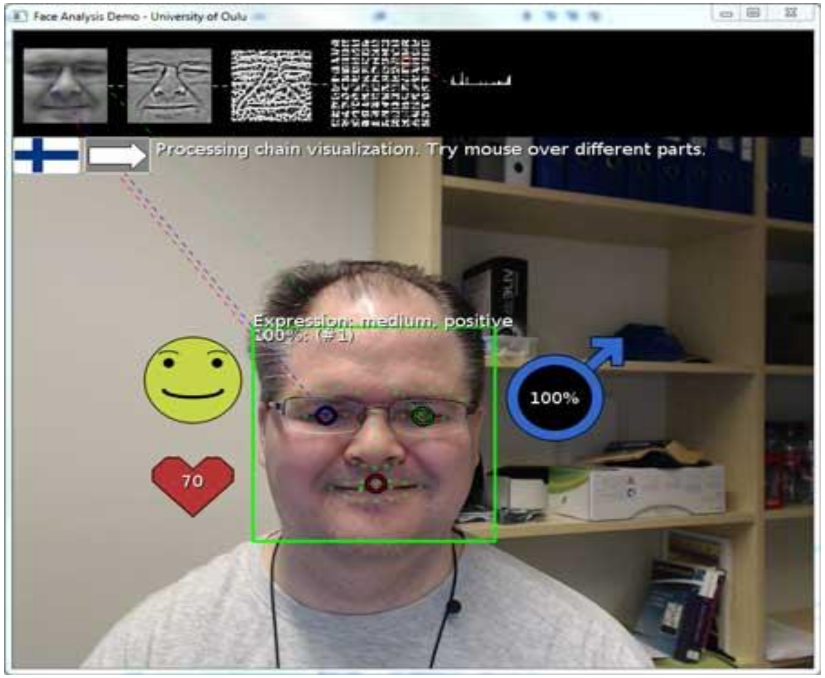
My 10 Year Research Vision: Face Analysis Systems

Matti Pietikäinen
Center for Machine Vision Research
University of Oulu, Finland
http://www.cse.oulu.fi/CMV



Face Analysis Demo - University of Oulu

Processing chain visualization. Try mouse over different parts.

Expression: medium, positive
100%: (#1)

Infotech Oulu

UNIVERSITY of OULU
OULUN YLIOPISTO

Human faces contain lots of information

- Identity
- Demographic information (e.g., gender, age, race/ethnicity ...)
- Emotions (”happy”, ”sad” ”angry”, ”surprise”, etc.)
- Direction of attention (head pose, gaze direction)
- Visual speech
- Health (e.g., pain, psychiatric diseases,...)
- Even ”unvisible” information (e.g., heart rate, micro-expressions)

Face information is very important, e.g. for perceptual interfaces in intelligent HCI

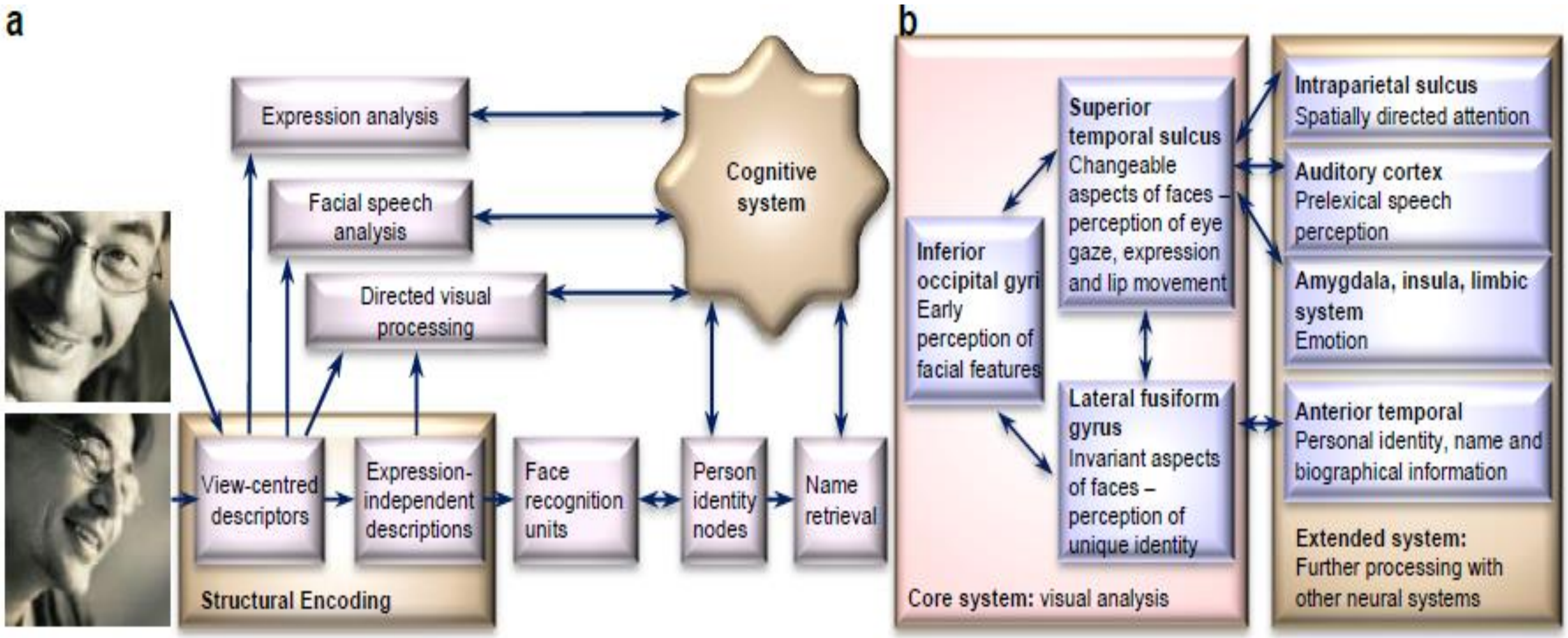
My vision

Past research on facial image analysis has focused on more or less isolated problems, including facial feature extraction, face description, face detection, face alignment, face recognition, facial expression recognition, pose and gaze estimation, soft biometrics, visual speech recognition.

My research vision is that many subareas of facial image analysis are mature enough – and more focus should be given to the problems of (multimodal) face analysis systems.

My grand challenge is to build a seeing and talking face for natural human-computer interaction.

Cognitive models of human perception may give some hints for computational systems, e.g. [Calder & Young, Nat Rev Neurosci, 2005]

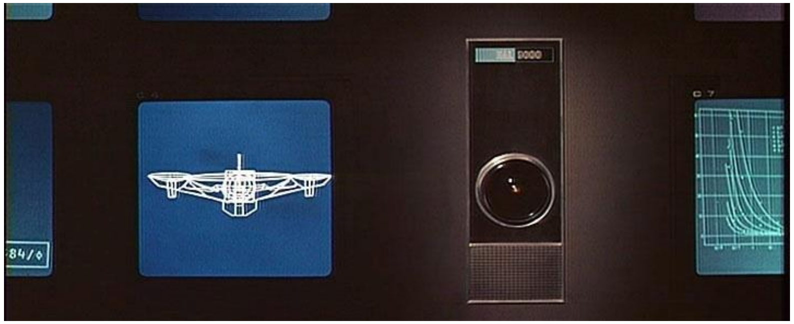

a
Expression analysis
Facial speech analysis
Directed visual processing
Cognitive system
View-centred descriptors
Expression-independent descriptions
Structural Encoding
Face recognition units
Person identity nodes
Name retrieval
b
Inferior occipital gyri
Early perception of facial features
Superior temporal sulcus
Changeable aspects of faces – perception of eye gaze, expression and lip movement
Lateral fusiform gyrus
Invariant aspects of faces – perception of unique identity
Core system: visual analysis
Intraparietal sulcus
Spatially directed attention
Auditory cortex
Prelexical speech perception
Amygdala, insula, limbic system
Emotion
Anterior temporal
Personal identity, name and biographical information
Extended system:
Further processing with other neural systems

Our future vision (from 2007): Human-centered ubiquitous systems - omnipresent, invisible, and imminent

- Technical wireless infrastructure everywhere in man-made environment, from wallpaper to vehicles to clothes to bloodflow
- Sensing, imaging, communications, ”intelligence” - and energy efficiency

- Machine vision will play a key role:
- Sensing and understanding human actions
- Face detection and recognition
- ”Lip reading”, gesture recognition
- Interpreting emotions

HAL 9000 on ”Discovery” spaceship was the first Ubicom system
(in Arthur C. Clarke’s / Stanley Kubrick’s 2001: A Space Odyssey; 1968)

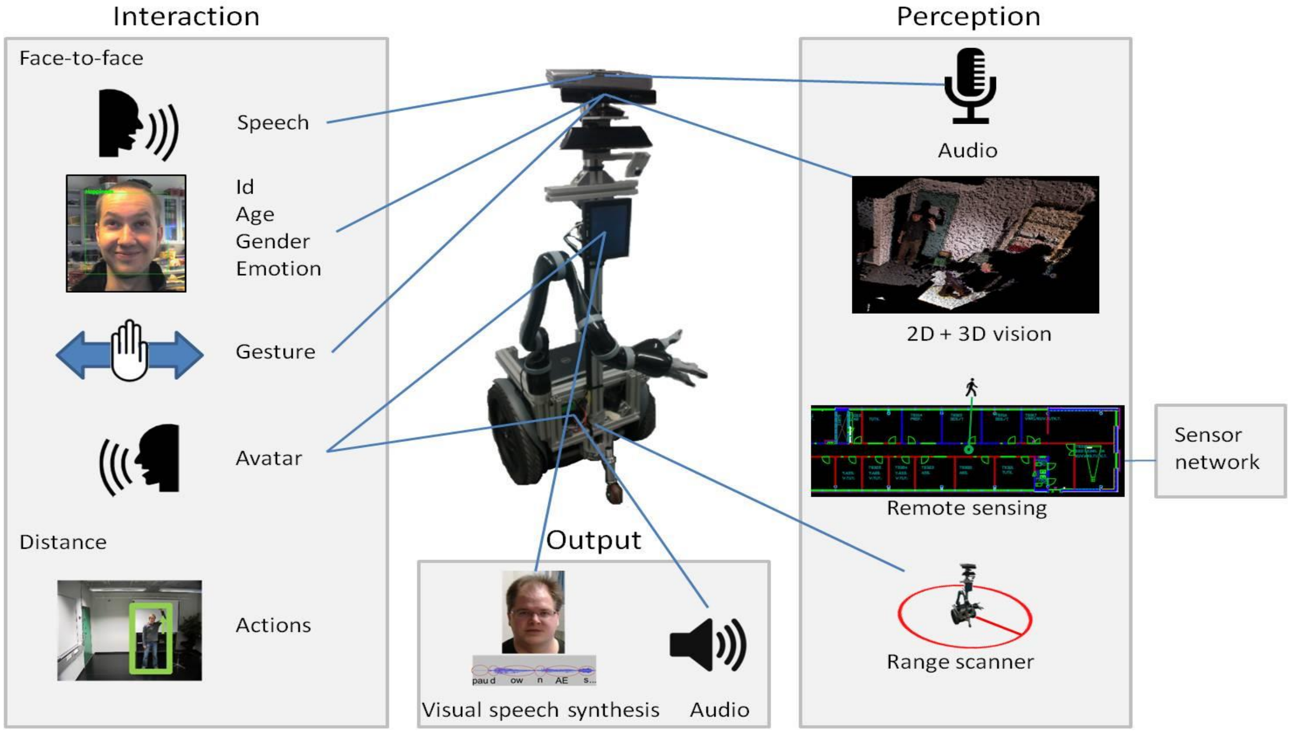


Example application: Human-robot interaction

- Interaction must be easy and natural

Example application: Interaction with a social robot

Röning J, Holappa J, Kellokumpu V, Tikanmäki A & Pietikäinen M (2014) Minotaurus: A system for affective human-robot interaction in smart environments. Cognitive Computation, 6(4):940-953.



Computer vision for human-computer interaction

- Development of natural, affective human-computer interfaces (HCI) is of great interest in building future ubiquitous computing systems

- We should be able to inteact with computers in a natural way, like in human-human interaction

- Computer vision will play a key role in building affective HCI systems

- The computer should be able to
  - detect and identify the user
  - recognize user’s emotions
  - communicate easily by understanding speech and gestures
  - provide a natural response based on its observation
  - detect and track humans; recognize their actions

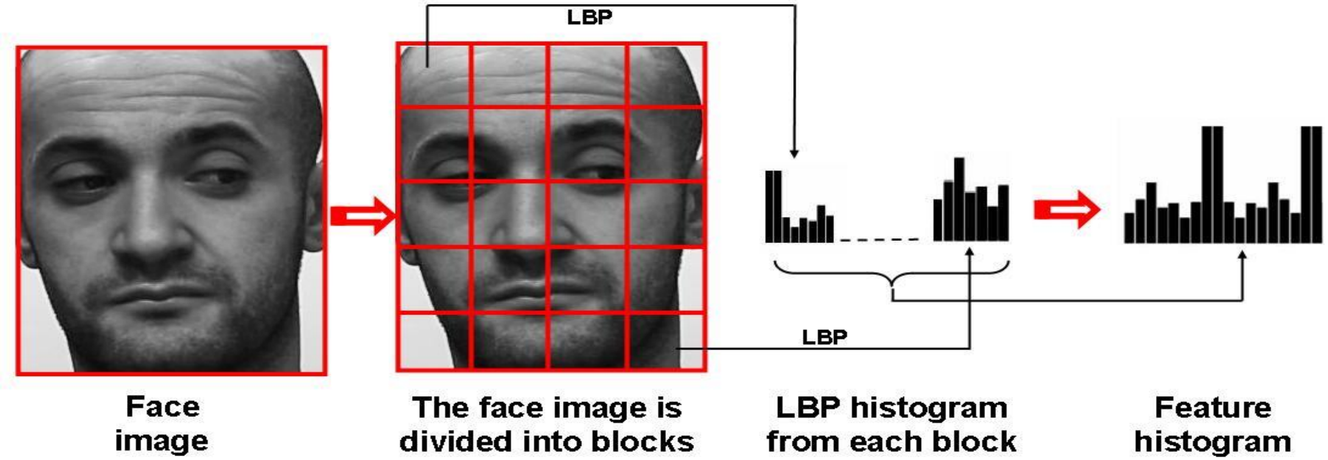
Towards natural HCI: Examples of key results from our recent research

- Face recognition and soft biometrics from near-frontal faces using LBPs
- Recognition of posed expressions using spatiotemporal LBPs
- Recognition of spontaneous facial microexpressions
- Recognition of spoken phrases from visual speech
- Remote heart rate measurement from video data
- 3D visual speech animation from video sequences

Face recognition with LBP

Ahonen T, Hadid A & Pietikäinen M (2006) Face description with local binary patterns: application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(12):2037-2041. (an early version published at ECCV 2004)

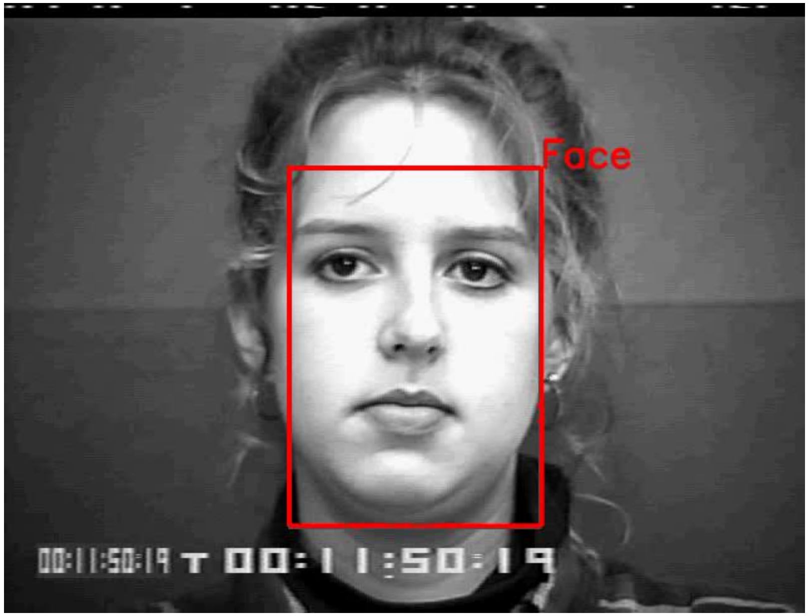
- ECCV paper was awarded **Koenderink Prize 2014** for fundamental contributions in computer vision


LBP
LBP
Face image
The face image is divided into blocks
LBP histogram from each block
Feature histogram

Facial expression recognition with LBP-TOP

Zhao G & Pietikäinen M (2007) Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(6):915-928.

- Introducing LBP-TOP operator
- Recognition of posed expressions from near-frontal faces



Recognition of spontaneous micro-expressions

Pfister T, Li X, Zhao G & Pietikäinen M (2011) Recognising spontaneous facial micro-expressions. Proc. International Conference on Computer Vision (ICCV 2011), Barcelona, Spain, 1449-1456.

What are micro-expressions?

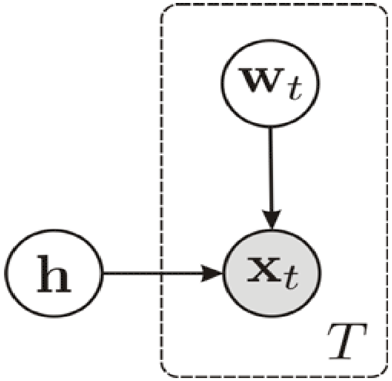
- Facial micro-expressions == rapid involuntary facial expressions which reveal suppressed affect (1/3 – 1/25 s)
- Currently only highly trained individuals are able to distinguish them, but even with proper training the recognition accuracy is only 47% (Frank 2009)

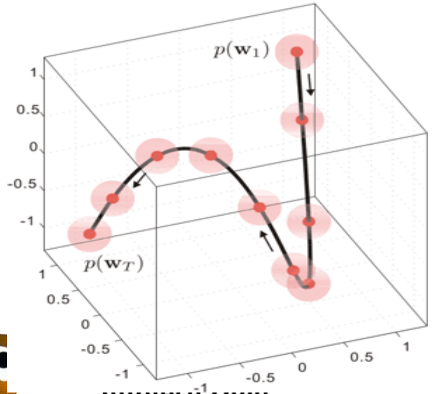
•1/10 speed

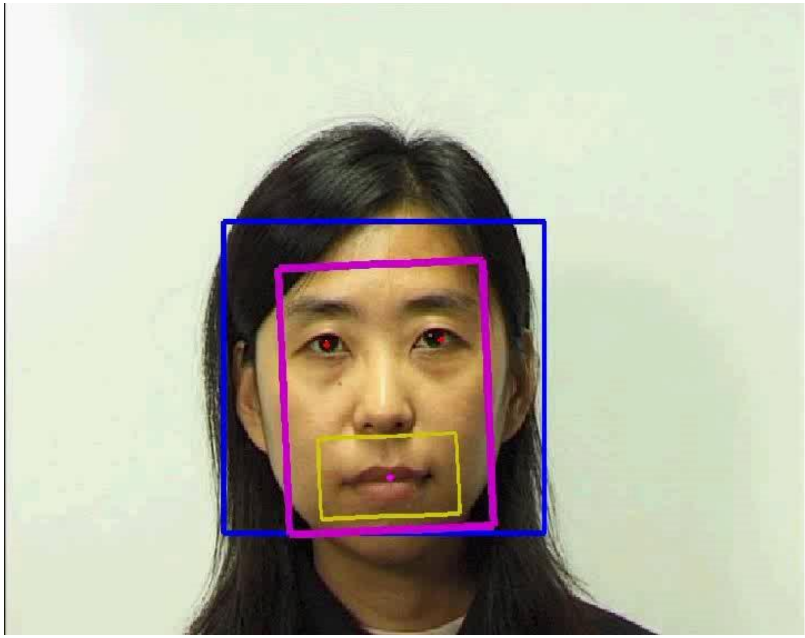
Visual speech recognition

Zhou Z, Hong X, Zhao G & Pietikäinen M (2014) A compact representation of visual speech data using latent variables. IEEE Transactions on Pattern Analysis and Machine Intelligence 36(1):181-187.

- Using generative latent variable model (GLVM) to model the inter-speaker variations of visual appearances and those caused by uttering

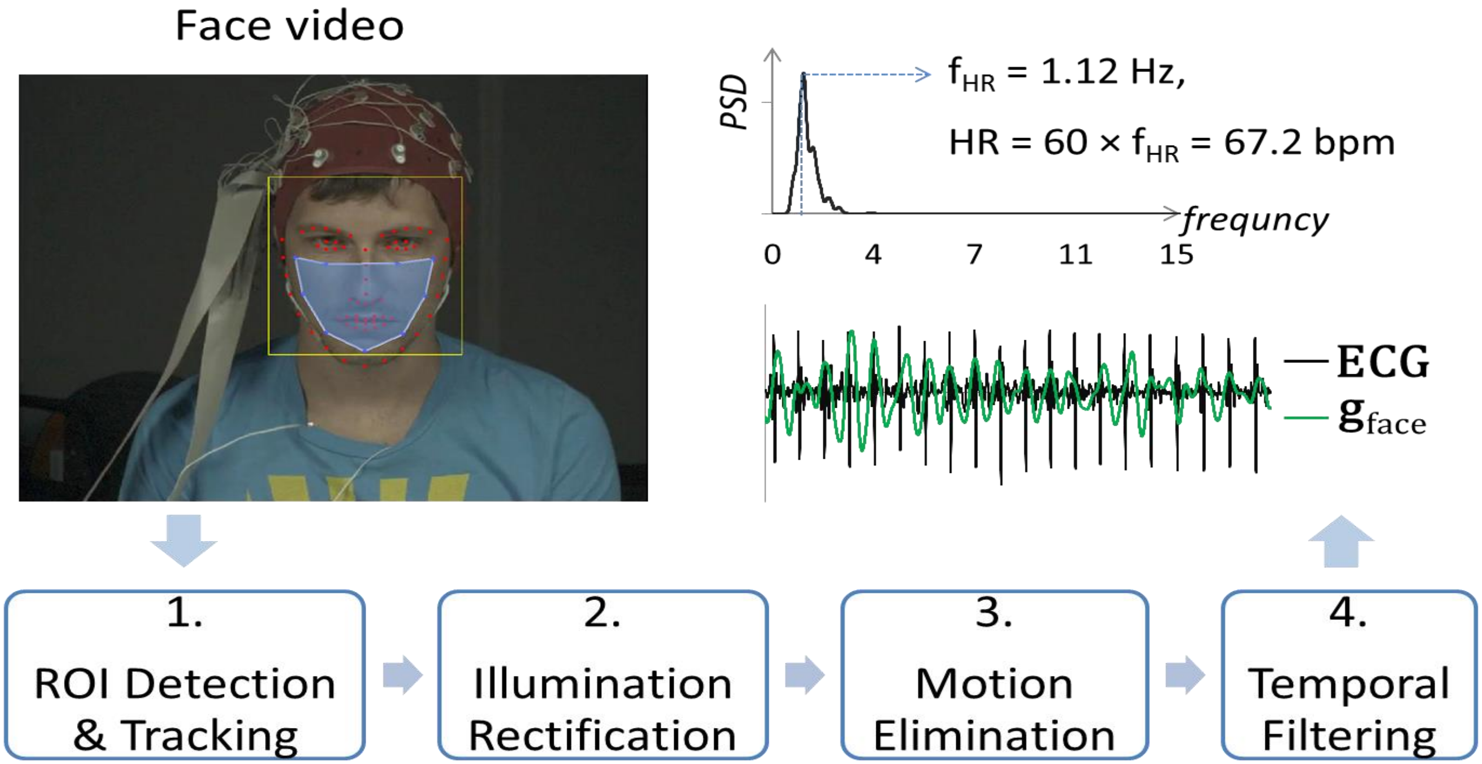






Heart rate from videos

Li X, Chen J, Zhao G & Pietikäinen M (2014) Remote heart rate measurement from face videos under realistic situations. Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014), Columbus, Ohio, 4265-4271.



Face video

f_{HR} = 1.12 Hz,

HR = 60 × f_{HR} = 67.2 bpm

1. ROI Detection & Tracking

2. Illumination Rectification

3. Motion Elimination

4. Temporal Filtering

Demo

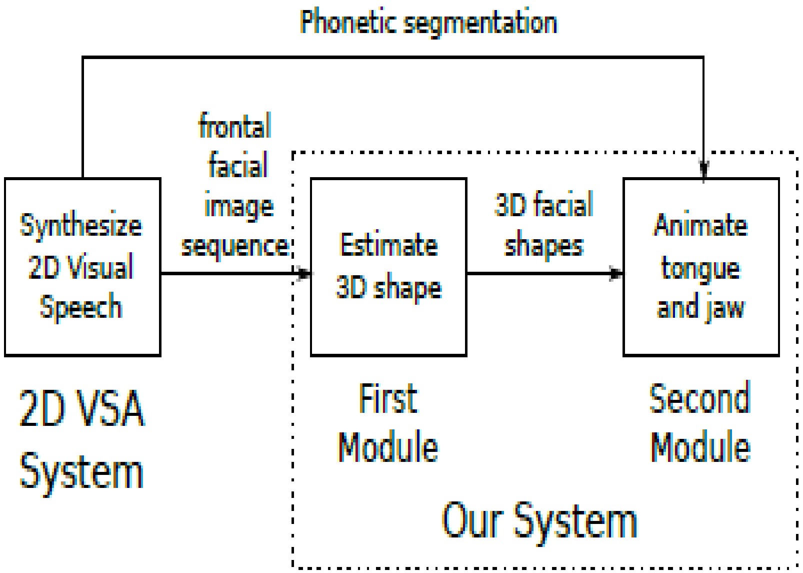
1st frame of input video

3D visual speech from video sequences

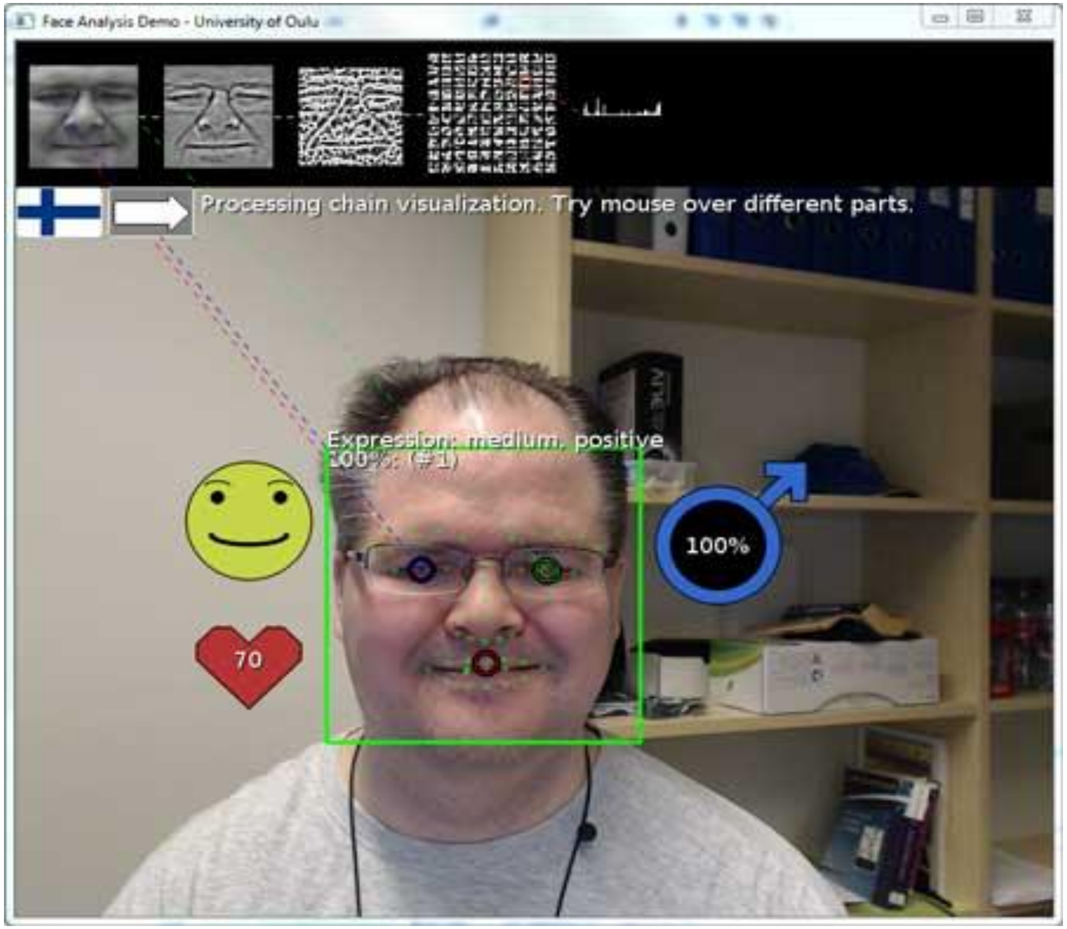
Musti U, Ouni S, Zhou Z & Pietikäinen M (2014) 3D visual speech animation from image sequences. Proc. The Ninth Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP 2014), Bangalore, India.

Motivation

- To synthesize 3D visual speech using a 2D Visual Speech Animation (VSA) system with a very small 3D visual speech corpus
- To have the advantage of a 2D VSA system in producing natural speech dynamics
- To have the renderability of a 3D VSA system.



A perceptual interface for face to face interaction



Processing chain visualization. Try mouse over different parts.

Expression: medium, positive
100%: (#1)

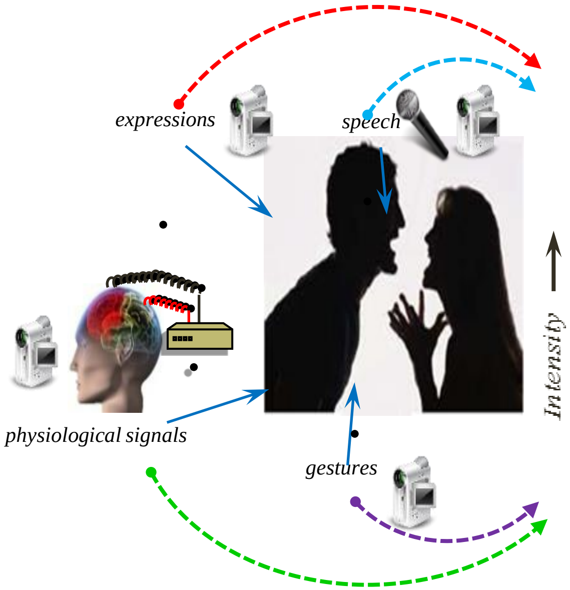
Challenges

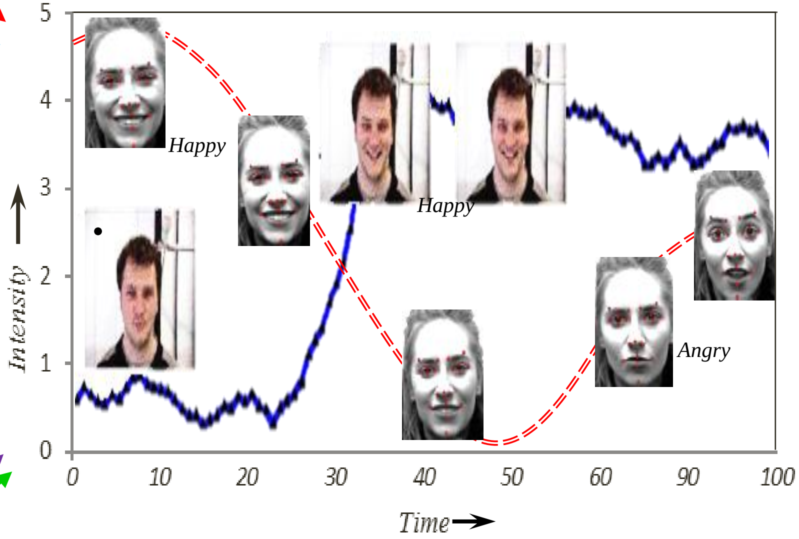
- Pose and illumination variations in real-world environments make all face analysis tasks much more difficult
- Recognition of spontaneous expressions is very hard
- Our understanding of visual speech is not yet as good as for audio speech
- Background noise makes audio speech recognition unreliable
- Training and testing of the system and its parts take lots of time and effort
- How to take into account interdependencies between different facial tasks?
- ...

Indentity recognition and soft biometrics

- Problems are caused by moving head with 3D pose and lighting variations, and talking face
- Robust descriptors needed for face analysis under changing conditions
- Soft-biometric information (e.g., gender, age, race/ethnicity) affects face recognition
- Facial expressions and audio-visual speech are useful for identity recognition
- More interaction, e.g. with speech recognition community would be useful

Recognition of spontaneous expressions

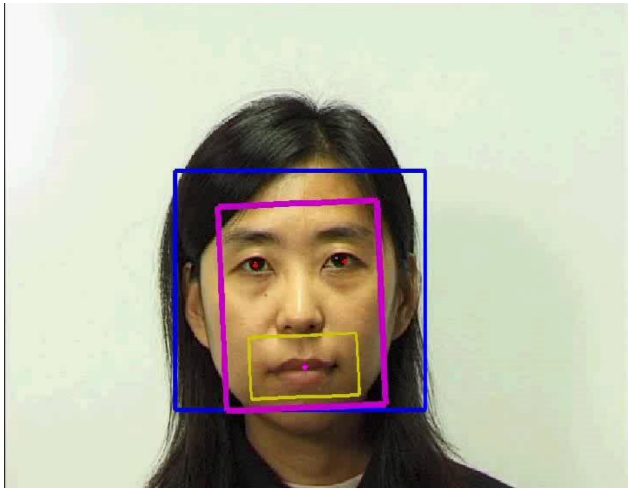




- Recognition of subtle spontaneous expressions from continuous video data is a great challenge
- Facial expressions are not culturally universal, also gender and facial age affect emotion decoding
- Joint use of different modalities (speech, head, body and eye movements, physiological signals (e.g., heart rate, temperature, respiration, galvanic skin response, blood pressure)
- Collaboration with experts from different disciplines, e.g., psychology, medicine

Audio-visual speech recognition

- Speech recognition in noisy conditions is very unreliable – a human listener can use visual cues (e.g., lip and tongue movements to enhance speech understanding)
- Gender, age and race/ethnicity affect speech recognition
- Our understanding of visual speech falls short of our understanding of the acoustic aspects of speech
- Recognizing continuous visual speech in real-world conditions would be vital for making ground-breaking progress in audio-visual speech recognition
- Collaboration with experts from different disciplines, e.g., audio speech, psychology

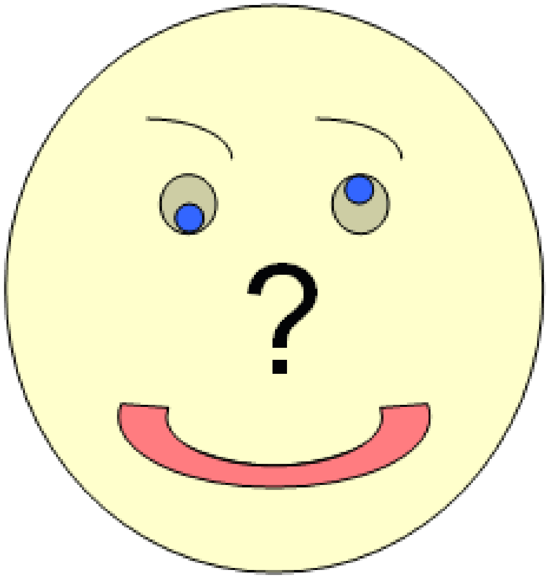


Infotech Oulu

UNIVERSITY of OULU
OULUN YLIOPISTO

Conclusions

- Human faces contain lots of useful information
- For face analysis systems multimodal information is often needed
- Face analysis systems should be considered as a whole, not as a collection of parts, because there are interdependencies between different parts
- More interaction between different disciplines would be needed (e.g., cognitive sciences, HCI, speech, graphics)
- Evaluation of multimodal face analysis systems is very challenging, requiring new test datasets and also system-level evaluation
- New applications, e.g. in wearable computing



Thanks!